

Annual report

VIAF Council meeting 2015

As an outgoing chair, I would like to set out and share with you some thoughts on the work led by the VIAF Council and highlight some forthcoming challenges faced by VIAF for the coming years, with an eye to good things and with the other eye to needed improvements. I shall precise that the following assessments are endorsed by the BnF.

To begin with, we need to characterize the VIAF Council functioning. As decided upon at the Singapore meeting in 2013, a conference call is held with the support of Linda Gabel (OCLC) about once every two months, with the participation of OCLC and of 5 to 7 VIAF contributors. It's a challenge for the coming year to have more participants, even if we are aware of the jet lag. Three main focuses of work can be identified.

1- The VIAFC produces policy documents. From the Lyon meeting on, the VIAF Council has been working on two policy documents which have been submitted for today's meeting: the adhesion criteria and the so-called "VIAF guidelines"). This later document was initially proposed by Ricardo Santos Munoz from the BNE.

2- Besides, the VIAF Council is a place for sharing practices which could have impacts on VIAF. For instance, as decided in Lyon, under the auspices of Brigitte Wiechmann from the DNB, a survey has been led on how the VIAF partners are implementing or are planning to implement the new MARC 21 fields on "country associated with a person" and on "professions".

3- The VIAF Council is of course a place for sharing information coming from OCLC on VIAF itself: evolutions of algorithms, the membership process for new applicants to join VIAF.

Some points could be improved. The first one is the issue of communication between participants, namely, for example, the use of the VIAF list serve. If a VIAF contributor has questions on the VIAF functioning, or information to share supposed to interest both OCLC and other VIAF contributors, either having possibly direct impacts on VIAF or falling within a wider area regarding authority data, it should not hesitate to use the VIAF list serve. It would be a good way for improving fluidity amongst VIAF partners. As you know, VIAF is very sensitive to evolutions of data handling, as they could have global impacts on VIAF itself and, consequently, on how data provided by others appear in VIAF clusters. For instance, if a partner decides to stop indicating birth dates for legal reasons, it not only implies that VIAF finds other ways to keep accuracy for merging records, but also it questions the presence of this birth date in data coming from other partners. A working group has begun working on the legal issues within the VIAFC two years ago. I think it would be helpful to resume its work in order to clarify how VIAF deals with them. It do not imply to have a shared legal policy on personal data, which is in the scope of national laws, but to think how to manage different legal frameworks in one data basis. And it is not easy.

The second point regards interrelations between VIAF contributors and OCLC. The VIAF algorithm is often considered by many VIAF end-users as a “black box”, that could be a blocking point to use VIAF. We know the VIAF algorithm is complex, depends on many factors, on intersecting rules and, first of all, quality of source data. From this point of view, some improvements can be noticed over the past few years. The monthly updates sent by Thom Hickey have informative, useful content. For instance, the move from Wikipedia to Wikidata last spring has been properly pre-announced and scheduled. The “VIAF guidelines” to be adopted will be helpful for VIAF members and applicants to pay attention to important MARC fields for VIAF algorithms. We can also greet the high responsiveness of technical OCLC team (Thom and Jenny Toves) when we see some problems. I have to point out that, based on such improvements, some progress remain necessary. For instance, we have learnt that, in the absence of other data, VIAF can merge some records by using only joint authors, which is quite unreliable. Besides, we have noticed OCLC decided to remove from VIAF names used as subjects, without warning *a priori* the VIAF Council. We had a good discussion with OCLC on this point, but only *after* the decision. So, I think that the VIAF Council is the right body to advice OCLC on planned evolutions of VIAF clustering, which implies to be informed about them. Moreover, the VIAF contributors are forced to submit proposals so as to improve alignments in VIAF. Many of them are indeed are using alignments processes in their own data and are gaining experience in such a field, for instance by implementing FRBRization processes or transition to RDA. Such process can be led with the OCLC support (for example: automatic creation of records on works by SUDOC). VIAF is both a precious tool for help them to achieve such plans, but also can draw inspiration from them to improve its own alignments. For instance, the BnF has developed a program for aligning manifestations and creating records for works in the project data.bnf.fr and the principles could be easily shared. It’s a point which is supposed to interest many libraries and VIAF itself with the background of FRBRization and transition to RDA.

Going on with transversal relationships amongst VIAF partners and between VIAF contributors and OCLC, it would be necessary to say a few words on the question of “quality”. For instance, if a given partners mixes two homonyms, it could have important consequences on other partners. The issue of quality remains a nodal question for VIAF. We know that the question of “quality” is very difficult in this time of budget and staff cuts. “Quality” doesn’t mean spending two hours for each record but implies to be clear on possible various quality levels and data sources (for instance; between records produced by libraries themselves and records retrieved from others, for instance from publishers). Both legal opportunities (open data) and technical ones (semantic web) encourage to reuse data coming from other partners, even beyond the library community. But, VIAF partners have to remain aware that, in all cases, an insufficient identification of a person could have huge consequences on data in VIAF. The [ISNI](#) quality team (up to now composed by BnF and BL) uses a list of addresses provided by VIAF partners in order to report them some errors. Our colleagues have an interesting and speed feedback from VIAF partners (15 VIAF partners have not yet given their contact), that shows a strong awareness of the quality issue from their part. It remains true that the issue of mixed homonyms in source data is a key-problem for VIAF. The VIAF partners and contributors are strongly encouraged to pay a particular attention on them.

The fourth point is related to strategic challenges faced by VIAF. As everybody knows, VIAF is at the crossroads of many key-evolutions and key-challenges, regarding authority data, FRBRization, semantic web, dialogues and data exchanges with other communities. The important, wide scope of reusing of data from VIAF is no more a topic to be debated. We are unable to list the projects using as a basis and as a core data coming from VIAF. The [IFLA satellite meeting on semantic web](#) held last year

in Paris gave the opportunity to VIAF to be cited almost 50 times on a single day. The workshop on VIAF in Strasbourg, under the auspices of OCLC EMEA annual meeting, in February 2013, and its huge attendance, is another example. These manifestations showed an interesting landscape of wide, large possibilities of reusing VIAF data and of the key-importance of VIAF in such a background. They also showed high expectations on evolutions of VIAF, regarding not only technical issues, but also the strategic positioning. It's also important to think about the next steps for VIAF having in mind the global data ecosystem and having in mind that the next evolutions of VIAF would have impacts on many projects and many libraries. It includes major national libraries and LAM agencies, which are the natural partners of OCLC, and international key-projects on persons, such as the ISO standard ISNI, also technically supported by OCLC¹. We could also cite other international projects, as [SNAC](#) (social network and archival context), under the aegis of NARA, which is designed to create authority data for the American archival network. We could also mention national projects, such as the National library of Poland's project to retrieve ISNI through VIAF for its digitalized library [Polona](#). When VIAF coughs, many key projects, either international, or national, are ill. VIAF now is of a crucial importance for libraries.

Currently, two databases maintained by OCLC fall in the scope of the names of persons and corporate bodies, VIAF and ISNI. The use of both VIAF and ISNI is spreading through the information community. Consequently, it's necessary to work on a global metadata strategy including both VIAF and ISNI. This strategy needs to be more specified and clarified. There is a high expectation among library leaders, given that many projects with high strategic and economic importance are depending either on VIAF or on ISNI, or both. Besides, more and more libraries are involved both in VIAF and in ISNI. We are aware that it's an important issue for OCLC too. The scope and aims of VIAF and ISNI are different. They do not overlap, as shown last year in the IFLA conference in Lyon. But, obviously, the links between VIAF and ISNI have a crucial importance. Again, it's not only a technical question, but also a strategic issue. This concerns an articulating strategic vision to be clarified and, this also could put the question of the business model of these services. I can stress libraries are eager to work with OCLC and to take part in the setting up a global strategic perspective for VIAF and ISNI. They can bring to OCLC their expertise on data production and data handling, as they are close to the source data. They can also build services of interest for OCLC. They could help and advice OCLC for working on a business model. Libraries are indeed eager to keep in mind the economic question.

Such considerations are beyond the scope of this VIAF Council meeting or, more precisely, can also be raised in other bodies but it's worth mentioning them here. The VIAF Council is an advisory body gathering some of the main OCLC partners, so it could kindly advice OCLC to have in mind such questions. They are, from our point of view, key issues over the coming years for improving the VIAF functioning and for making VIAF more sustainable in the long run. OCLC and libraries are in the same boat and I think that every one of us is willing to develop a win-win solution to face many common challenges.

¹ OCLC maintains and administers the ISNI central database, on behalf of the ISNI Registration Authority, which is governed by the ISNI-IA board.